Microsoft Academic Graph: When experts are not enough

Abstract

Abstract An ongoing project explores the extent to which artificial intelligence (AI), specifically in the areas of natural language processing and semantic reasoning, can be exploited to facilitate the studies of science by deploying software agents equipped with natural language understanding capabilities to read scholarly publications on the web. The knowledge extracted by these AI agents is organized into a heterogeneous graph, called Microsoft Academic Graph (MAG), where the nodes and the edges represent the entities engaging in scholarly communications and the relationships among them, respectively. The frequently updated data set and a few software tools central to the underlying AI components are distributed under an open data license for research and commercial applications. This paper describes the design, schema, and technical and business motivations behind MAG and elaborates how MAG can be used in analytics, search, and recommendation scenarios. How AI plays an important role in avoiding various biases and human induced errors in other data sets and how the technologies can be further improved in the future are also discussed.